SuperTarget goes quantitative: update on drug–target interactions

Nikolai Hecker[1], Jessica Ahmed[1], Joachim von Eichborn[1], Mathias Dunkel[1], Karel Macha[1], Andreas Eckert[1], Michael K. Gilson[2], Philip E. Bourne[2] and Robert Preissner[1,*]

[1]Structural Bioinformatics Group, Institute for Physiology, Charité-University Medicine Berlin, Lindenberger Weg 80, 13125 Berlin, Germany and [2]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, 9500 Gilman Drive, La Jolla CA 92093, USA

ABSTRACT

There are at least two good reasons for the on-going interest in drug–target interactions: first, drug-effects can only be fully understood by considering a complex network of interactions to multiple targets (so-called off-target effects) including metabolic and signaling pathways; second, it is crucial to consider drug-target-pathway relations for the identification of novel targets for drug development. To address this on-going need, we have developed a web-based data warehouse named *SuperTarget*, which integrates drug-related information associated with medical indications, adverse drug effects, drug metabolism, pathways and Gene Ontology (GO) terms for target proteins. At present, the updated database contains >6000 target proteins, which are annotated with >330 000 relations to 196 000 compounds (including approved drugs); the vast majority of interactions include binding affinities and pointers to the respective literature sources. The user interface provides tools for drug screening and target similarity inclusion. A query interface enables the user to pose complex queries, for example, to find drugs that target a certain pathway, interacting drugs that are metabolized by the same cytochrome P450 or drugs that target proteins within a certain affinity range. *SuperTarget* is available at http://bioinformatics.charite.de/supertarget.

INTRODUCTION

In the last decade, non-commercial drug- or target-related databases have been established. Millions of compounds can be found in databases like ChEMBL (1) or PubChem (2) and their availability can be checked via databases like ZINC (3).

Several databases collect binding data on small molecules, in particular, drugs. A comprehensive and manually curated resource is DrugBank (4), which contains 4300 targets related to about 7000 compounds, including 1500 FDA-approved drugs. Another notable database is the Therapeutic Target Database (TTD) (5), which holds target information on approximately 2000 classified targets linked to 5000 compounds, including 1500 approved drugs. Surprisingly, the overlap between these two databases is small (data not shown). KEGG DRUG is a database for approved drugs, which comprises drug–target interactions, drug classifications as well as information about drug structure development (6). Other databases collect drug–target data with a special focus regarding medical indications [e.g. cancer (7) and infection (8)], technical aspects [e.g. pharmacophores (9) or scaffold hoppers (10)], side effects (11) or special metabolic pathways (12). The database STITCH is focused on the relation of >70 000 chemicals to targets from hundreds of different organisms (13). To understand the complex effects of drugs, the relation of their targets in signaling and metabolic pathways are important and reflected in a number of databases, e.g. KEGG (6) or Reactome (14).

In 2008, the first version of *SuperTarget* was developed with the intention to accentuate drug–target interactions themselves and to provide references to other resources for more elaborate analysis (15). Adverse drug reactions are a common reason for the rejection of drug candidates during clinical trial or withdrawal after approval. For example, the cyclo-oxygenase inhibitor rofecoxib (nonsteroidal anti-inflammatory drug) was withdrawn worldwide because of severe cardiovascular side effects, which may be caused by unanticipated interactions with potassium and calcium channels (16).

The analysis of drug–target interactions can play a crucial role to improve the process of drug design and

*To whom correspondence should be addressed. Tel: +49 30 450540755; Fax: +49 30 450540955; Email: robert.preissner@charite.de

admission. *SuperTarget* provided a variety of drug–target interactions and affected biological pathways in a user-friendly manner. This second release of *SuperTarget* contains a core dataset of ∼330 000 drug–target interactions, of which about 310 000 interactions have binding affinity data. We consider a drug–target relation as a specific interaction of a small chemical compound, which could be used to treat or diagnose a disease. Thus, *SuperTarget* now enables scientists to carry out not only qualitative but also quantitative analysis of drug–target interactions.

DATA SET

SuperTarget at present contains an updated version of the original dataset. In 2011, the core dataset consists of 6219 targets and 195 770 drugs and putative drugs of which about 2500 are approved drugs, that are classified by the World Health Organization (WHO), resulting in 332 828 drug–target interactions. The list of targets was selected using the PROMISCUOUS database (17). New drug interactions were added from inhouse text mining, supplemented by manual curations, SuperSite (18), CaRe (7), SuperCyp (12) and DrugBank (4) using the target list mentioned above. Target synonyms and external database identifiers were updated as defined in the UniProtKB database (19). The information content of drug entities was enlarged by general properties such as molecular weight, lipophilicity (logP) and known side effects as defined by the Sider database (11). Binding affinities of drug–target interactions were added from BindingDB entries (20). *SuperTarget* offers references to various resources, which provide more detailed information, i.e. specific links to PubChem, UniProtKB, DrugBank, the RCSB Protein Data Bank (PDB), PubMed and the BindingDB.

Relations from other databases, namely DrugBank (4), KEGG (6), PDB(21), SuperDrug (22) and TTD (5) were checked for drug–target interactions not identified using the preceding steps. If those interactions could be confirmed by literature listed in PubMed, the references were included in *SuperTarget* otherwise the describing database is referenced. To provide users with further information on drug–target interactions, *SuperTarget* provides links to physicochemical properties and further structural information of drugs. Proven or potential target proteins are represented as stored in UniProtKB (19), by functional annotations extracted from GO (23), and by related pathway information provided by KEGG (6) (compare Figure 1).

FEATURES

SuperTarget enables users to link drugs and targets to biomolecular pathways. Pathways are given as defined in the KEGG database. Furthermore, targets can be searched using gene ontology (GO) terms. The Anatomical Therapeutical Chemical (ATC) classification of drugs (24) is useful for searching drugs in distinct indication areas and for analyzing co-occurrence of drugs/targets in

different diseases, which could be combined with side effect searches. A special section is dedicated to drug interactions with enzymes of the Cytochrome P450 family (CYP). CYPs are mono-oxygenases whose functions in humans include the detoxification of foreign substances via chemical modification. Hence, they play a crucial role in drug metabolism. The features described above were part of the first release, but have now been updated extensively. Among other new features, *SuperTarget* now introduces integration of protein–protein interaction data from the ConsensusPathDB (25). In addition, target sequence and drug similarities are incorporated in *SuperTarget*. Drug similarities are computed using Tanimoto coefficients of pre-calculated fingerprints as implemented in SuperPred (26).

WEB INTERFACE

The accessibility and presentation of data is as important as its accumulation and integration. The *SuperTarget* web interface offers a variety of ways to obtain and view information about the drugs and targets:

 (i) To provide quick access to the data, a simple full-text search called 'Targle' was implemented, which returns hits separately for drugs, targets and pathways.
 (ii) There is also a dedicated search section for each type of entity: i.e. drugs, targets, pathways, gene ontologies and CYPs. The user is either able to select predefined identifiers or to type in a variety of search terms. For instance, targets can be searched by synonyms, UniProtKB identifiers, PDB identifiers, KEGG target identifiers or EC numbers. The results section for each entity provides detailed information about each instance, which includes SMILES and InChI strings and a list of putative targets for drug results as well as a list of protein–protein interactions and similar targets for target results. All different entities are cross-linked, thus details of putative targets and affected pathways are easily accessible from drug search results.
(iii) For more sophisticated searches, an advanced search option is available, which includes general properties, e.g. a desired number of H-bond donors, or characteristics associated with particular drugs or targets such as affinity values. This option allows an arbitrary combination of different search criteria. Hence, the user is able to perform a variety of complex searches.

CASE STUDY: VASCULAR ENDOTHELIAL GROWTH FACTOR RECEPTORS

The following case study illustrates a useful application of the advanced search capability. Pathways, which include Vascular Endothelial Growth Factors (VEGF), play an important role in angiogenesis and provide targets for anti-angiogenic cancer therapy (27). It has been shown that normalization of tumor vasculature via inhibition of
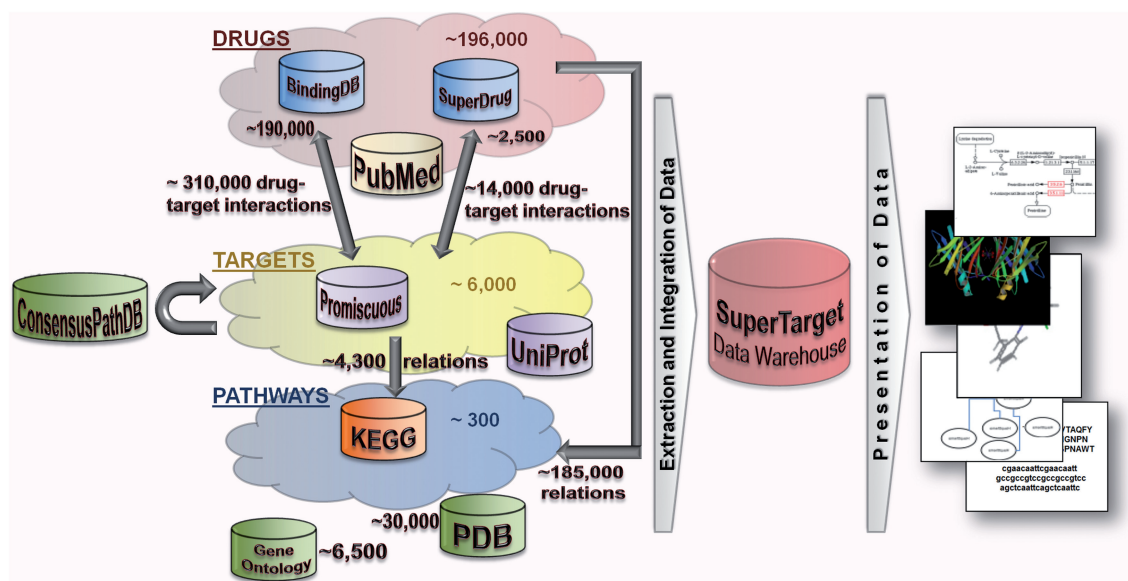
Figure 1. System architecture and number of database entries of *SuperTarget*. The three clouds represent drugs (red), targets (yellow) and pathways (blue) with given numbers, while the numbers of the respective relations are given at the connecting arrows. Beside the targets, drugs and pathways the database provides >6500 GO-terms associated with drug–target interaction and 30 000 links to protein structures are given. Furthermore, protein–protein interactions from the ConsensusPathDB are included.

VEGFR-2 receptor reduces hypoxia and cell proliferation and increases accessibility to other chemotherapeutic drugs (28). Therefore, it may be desirable to find specific inhibitors of VEGFR-2, which do not affect the other receptor subtype VEGFR-1. A general idea about the specificity of an inhibitor is provided by its IC50 value. This refers to the compound concentration, which is needed for a half-maximum inhibition of the corresponding biological process. Thus, potent inhibitors show low IC50 values whereas weak inhibitors exhibit relatively high ones. The IC50 values for the two different VEGFR subtypes should differ by at least an order of magnitude (e.g. nM versus μM). Compounds with the desired properties can be searched in *SuperTarget* as follows:

First, both receptor subtypes are searched by their UniProt name in the target section and are added to the basket. Second, the advanced search section is used to identify potential drugs that inhibit VEGFR-2 but not VEGFR-1. In more detail, the VEGFR-1 receptor subtype is added from the basket to the query as an 'is ligand of ("VGFR1_HUMAN")' search criterion. A range of IC50 values is defined as a second search criterion (1001–100 000 nM for VEGFR-1). This criterion is automatically associated with the first one. The steps mentioned above are then repeated for the second receptor subtype, this time the desired range for the IC50 value is 0–100 nM. This query returns nine candidates. Since different experiments may suggest different IC50 values for the same biological process, the results have to be manually curated. The corresponding information is shown in the interaction detail sections of the compounds and VEGFR-1 or VEGFR-2, which are accessible from the drug details page of each compound. The results of our

analysis suggest that at least six putative drugs with this feature may be worth further investigation, i.e. ChEMBL205413, ChEMBL205610, ChEMBL209919, ChEMBL213507, ChEMBL380397 and ChEMBL398610 (Figure 2).

In two following steps, the crude cell biological context of these six drugs can be analyzed. The six putative drugs are added to the basket. In the next step, additional targets of the six putative drugs can be identified. Similarly to the steps mentioned above, for each of the six putative drugs the advanced search option is used to search targets, which are strongly inhibited by the compound, i.e. have IC50 values between 0 and 100 nM. Each additional target is added to the basket. Three of the putative drugs, ChEMBL209919, ChEMBL213507 and ChEMBL380397, also show a strong inhibitory effect on the mast/stem cell-growth factor receptor (UniProt name: FLT3_HUMAN). ChEMBL398610 exhibits low IC50 values toward the FL cytokine receptor (UniProt name: FLT3_HUMAN), the high affinity nerve growth factor receptor (UniProt name: NTRK1_HUMAN) and VEGFR-3 (UniProt name: VGFR3_HUMAN).

In a final step, the list of additional targets can be used to identify pathways, which are affected by the six putative drugs mentioned in the first paragraph of this case study. Either the advanced search option or the pathway link in the list of results can be used to identify pathways containing VEGFR-2. VEGFR-2 is associated with the human cytokine–cytokine interaction, endocytosis, focal adhesion and the VEGF signaling pathway. All of the six putative drugs are likely to have an impact on these pathways. In a similar fashion, additional pathways comprising the mast/stem cell growth factor receptor, which is a target of ChEMBL209919, ChEMBL213507
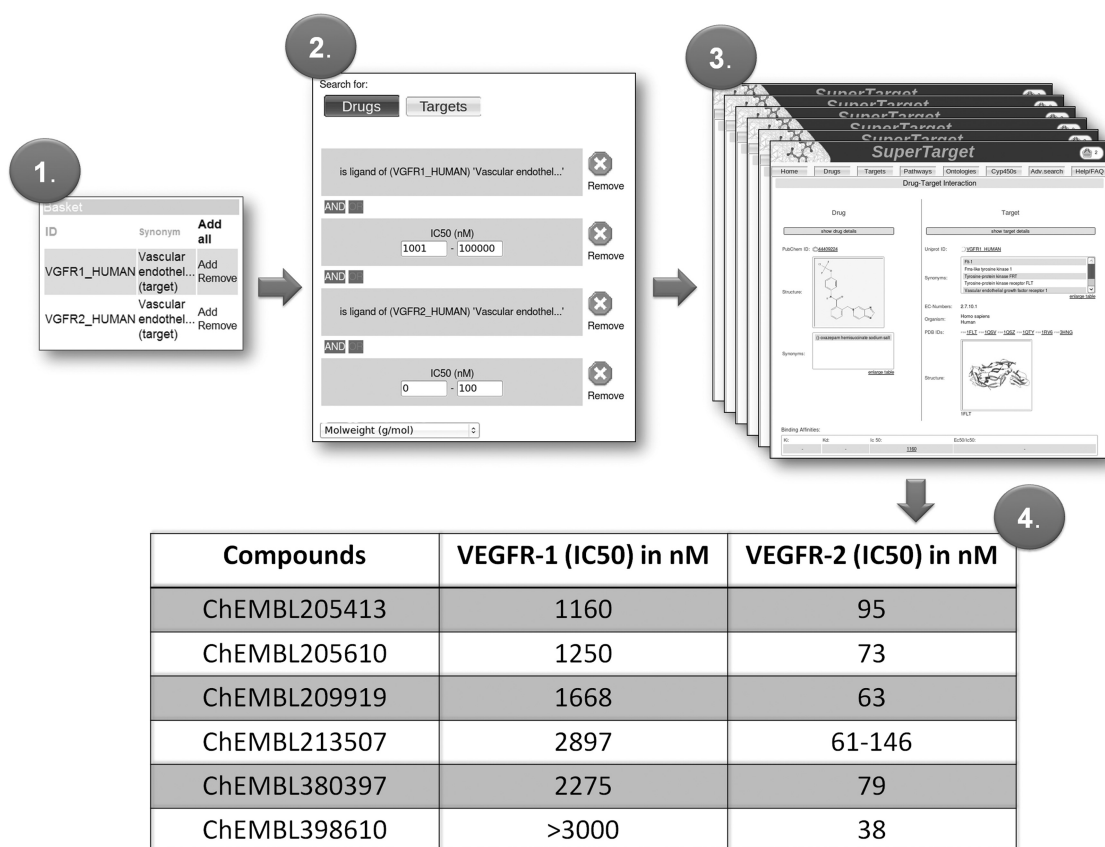
Compounds	VEGFR-1 (IC50) in nM	VEGFR-2 (IC50) in nM
ChEMBL205413	1160	95
ChEMBL205610	1250	73
ChEMBL209919	1668	63
ChEMBL213507	2897	61-146
ChEMBL380397	2275	79
ChEMBL398610	>3000	38

Figure 2. Case study: query and results. (1) Target search results for VEGFR-1 and VEGFR-2 are added to the basket. (2) Search criteria are defined: select all drugs, which showed a low VEGFR-1 binding affinity (IC50 value between 1001 and 100 000) but a high VEGFR-2 affinity (IC50 value between 0 and 100 nM). (3) Result pages for each drug–target relation show detailed information on drug, target and binding affinity and (4) An examination of the query results identifies six compounds with the desired properties. Refer to the text for further information.

and ChEMBL380397, are the human hemapoietic cell lineage, melagonesis, acute myeloid leukemia pathway and pathways in cancer. For the compound ChEMBL398610, the union of pathways, which contain at least one of its targets, appears to be interesting, i.e. pathways comprising the FL cytokine receptor or the high-affinity nerve growth factor receptor or VEGFR-3. Hence, these three targets are added as pathway search criteria in the advanced search section and combined by OR conjunctions. In addition to the pathways associated with VEGFR-2, the results include the Human MAPK signaling, neurotrophin signaling, hemapoietic cell lineage, apoptosis, acute myeloid leukemia, thyroid cancer pathway and pathways in cancer.

CONCLUSION

SuperTarget is one of the largest resources of validated drug–target interactions including quantitative data. We carefully assessed the most relevant data sections for each entity and provide links to many other resources. The updated search engine allows users to obtain information starting from a variety of entry points and to perform complex queries.

AVAILABILITY

SuperTarget is available via the web site without registration: http://bioinformatics.charite.de/supertarget. It can be used under the terms of the Creative Commons Attribution-Noncommercial-Share Alike 3.0 License.

Conflict of interest statement. None declared.

REFERENCES

1. Gaulton,A., Bellis,L.J., Bento,A.P., Chambers,J., Davies,M., Hersey,A., Light,Y., McGlinchey,S., Michalovich,D.,

Al-Lazikani,B. *et al.* (2012) ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.*, **40**, D1100–D1107.

2. Wang,Y., Xiao,J., Suzek,T.O., Zhang,J., Wang,J. and Bryant,S.H. (2009) PubChem: a public information system for analyzing bioactivities of small molecules. *Nucleic Acids Res.*, **37**, W623–633.

3. Irwin,J.J. and Shoichet,B.K. (2005) ZINC—a free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model*, **45**, 177–182.

4. Knox,C., Law,V., Jewison,T., Liu,P., Ly,S., Frolkis,A., Pon,A., Banco,K., Mak,C., Neveu,V. *et al.* (2011) DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. *Nucleic Acids Res.*, **39**, D1035–1041.

5. Zhu,F., Han,B., Kumar,P., Liu,X., Ma,X., Wei,X., Huang,L., Guo,Y., Han,L., Zheng,C. *et al.* (2010) Update of TTD: Therapeutic Target Database. *Nucleic Acids Res.*, **38**, D787–791.

6. Kanehisa,M., Goto,S., Furumichi,M., Tanabe,M. and Hirakawa,M. (2010) KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.*, **38**, D355–360.

7. Ahmed,J., Meinel,T., Dunkel,M., Murgueitio,M.S., Adams,R., Blasse,C., Eckert,A., Preissner,S. and Preissner,R. (2011) CancerResource: a comprehensive database of cancer-relevant proteins and compound interactions supported by experimental knowledge. *Nucleic Acids Res.*, **39**, D960–967.

8. Barh,D., Kumar,A. and Misra,A.N. (2010) Genomic Target Database (GTD): a database of potential targets in human pathogenic bacteria. *Bioinformation*, **4**, 50–51.

9. Hsin,K.Y., Morgan,H.P., Shave,S.R., Hinton,A.C., Taylor,P. and Walkinshaw,M.D. (2011) EDULISS: a small-molecule database with data-mining and pharmacophore searching capabilities. *Nucleic Acids Res.*, **39**, D1042–1048.

10. Yan,B.B., Xue,M.Z., Xiong,B., Liu,K., Hu,D.Y. and Shen,J.K. (2009) ScafBank: a public comprehensive scaffold database to support molecular hopping. *Acta Pharmacol. Sin.*, **30**, 251–258.

11. Kuhn,M., Campillos,M., Letunic,I., Jensen,L.J. and Bork,P. (2010) A side effect resource to capture phenotypic effects of drugs. *Mol. Syst. Biol.*, **6**, 343.

12. Preissner,S., Kroll,K., Dunkel,M., Senger,C., Goldsobel,G., Kuzman,D., Guenther,S., Winnenburg,R., Schroeder,M. and Preissner,R. (2010) SuperCYP: a comprehensive database on Cytochrome P450 enzymes including a tool for analysis of CYP-drug interactions. *Nucleic Acids Res.*, **38**, D237–243.

13. Kuhn,M., Szklarczyk,D., Franceschini,A., Campillos,M., von Mering,C., Jensen,L.J., Beyer,A. and Bork,P. (2010) STITCH 2: an interaction network database for small molecules and proteins. *Nucleic Acids Res.*, **38**, D552–556.

14. Croft,D., O'Kelly,G., Wu,G., Haw,R., Gillespie,M., Matthews,L., Caudy,M., Garapati,P., Gopinath,G., Jassal,B. *et al.* (2011) Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res.*, **39**, D691–697.

15. Gunther,S., Kuhn,M., Dunkel,M., Campillos,M., Senger,C., Petsalaki,E., Ahmed,J., Urdiales,E.G., Gewiess,A., Jensen,L.J. *et al.* (2008) SuperTarget and Matador: resources for exploring drug–target relationships. *Nucleic Acids Res.*, **36**, D919–922.

16. Brueggemann,L.I., Mani,B.K., Mackie,A.R., Cribbs,L.L. and Byron,K.L. (2010) Novel actions of nonsteroidal anti-inflammatory drugs on vascular ion channels: accounting for cardiovascular side effects and identifying new therapeutic applications. *Mol. Cell. Pharmacol.*, **2**, 15–19.

17. von Eichborn,J., Murgueitio,M.S., Dunkel,M., Koerner,S., Bourne,P.E. and Preissner,R. (2011) PROMISCUOUS: a database for network-based drug-repositioning. *Nucleic Acids Res.*, **39**, D1060–1066.

18. Bauer,R.A., Gunther,S., Jansen,D., Heeger,C., Thaben,P.F. and Preissner,R. (2009) SuperSite: dictionary of metabolite and drug binding sites in proteins. *Nucleic Acids Res.*, **37**, D195–200.

19. UniProt Consortium. (2010) The Universal Protein Resource (UniProt) in 2010. *Nucleic Acids Res.*, **38**, D142–148.

20. Liu,T., Lin,Y., Wen,X., Jorissen,R.N. and Gilson,M.K. (2007) BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities. *Nucleic Acids Res.*, **35**, D198–201.

21. Rose,P.W., Beran,B., Bi,C., Bluhm,W.F., Dimitropoulos,D., Goodsell,D.S., Prlic,A., Quesada,M., Quinn,G.B., Westbrook,J.D. *et al.* (2011) The RCSB Protein Data Bank: redesigned web site and web services. *Nucleic Acids Res.*, **39**, D392–401.

22. Goede,A., Dunkel,M., Mester,N., Frommel,C. and Preissner,R. (2005) SuperDrug: a conformational drug database. *Bioinformatics*, **21**, 1751–1753.

23. Barrell,D., Dimmer,E., Huntley,R.P., Binns,D., O'Donovan,C. and Apweiler,R. (2009) The GOA database in 2009—an integrated Gene Ontology Annotation resource. *Nucleic Acids Res.*, **37**, D396–403.

24. WHO Expert Committee. (2009) The selection and use of essential medicines. *World Health Organ Tech Rep Ser*, 1–242, back cover.

25. Kamburov,A., Pentchev,K., Galicka,H., Wierling,C., Lehrach,H. and Herwig,R. (2011) ConsensusPathDB: toward a more complete picture of cell biology. *Nucleic Acids Res.*, **39**, D712–717.

26. Dunkel,M., Gunther,S., Ahmed,J., Wittig,B. and Preissner,R. (2008) SuperPred: drug classification and target prediction. *Nucleic Acids Res.*, **36**, W55–59.

27. Olsson,A.K., Dimberg,A., Kreuger,J. and Claesson-Welsh,L. (2006) VEGF receptor signalling - in control of vascular function. *Nat. Rev. Mol. Cell Biol.*, **7**, 359–371.

28. Jain,R.K. (2005) Normalization of tumor vasculature: an emerging concept in antiangiogenic therapy. *Science*, **307**, 58–62.